RPcalc User's Guide (Version 2.0)

A program to determine when effluent limitations are needed

April 26, 2005

Steven G. Saiz, Environmental Scientist

California Environmental Protection Agency
State Water Resources Control Board
Division of Water Quality
1001 I St., Sacramento, CA 95814
(916) 341-5582
ssaiz@waterboards.ca.gov

Introduction

This computer program will compare an environmental data set with a fixed regulatory standard. This program was developed to assist permit writers when conducting a "reasonable potential" analysis for ocean discharges in accordance with the California Ocean Plan (SWRCB 2001). More details on this procedure may be found in Issue 1 of the Final Functional Equivalent Document of the Water Quality Control Plan for Ocean Waters of California (SWRCB 2005).

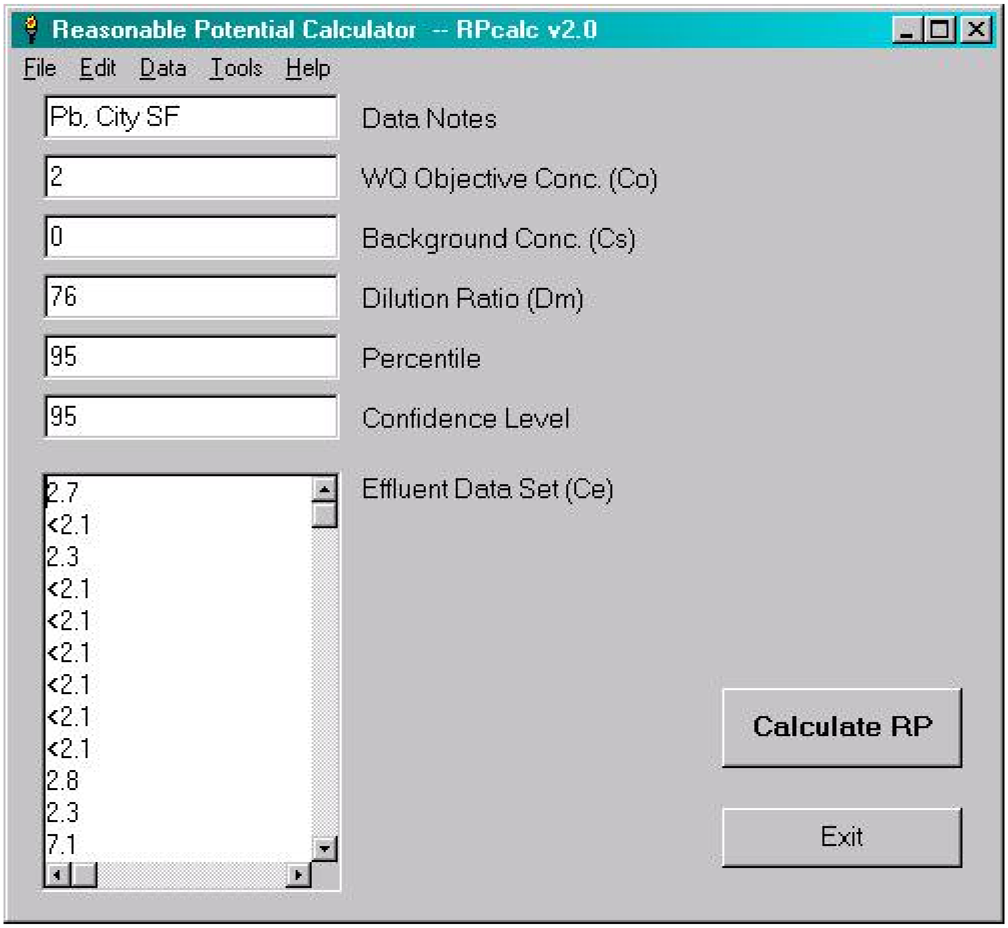
Whenever possible, this program will calculate the one-sided, upper confidence bound (UCB) of an effluent population percentile after complete mixing. The UCB is compared to a numeric Water Quality Objective or criterion to determine whether an effluent limitation is required under the National Pollutant Discharge Elimination System (NPDES) regulations, 40 CFR 122.44, i.e., whether the discharge causes, has the "reasonable potential" to cause, or contributes to an excursion above the state Water Quality Objective.

If the UCB does not exceed the Water Quality Objective then an effluent limitation for the pollutant is not usually required. Conversely, if the UCB or any measured sample value exceeds the Water Quality Objective then there is a reasonable potential to exceed the Water Quality Objective and an NPDES effluent limitation is required.

RPcalc accounts for effluent dilution and background pollutant levels by calculating the UCB after complete mixing in the regulatory mixing zone. This program will accept multiply censored effluent concentration data and analyzes this "censored" data set using the robust probability plotting method of Helsel & Cohn (1988). More extreme data censoring leads to a non-parametric reasonable potential data analysis.

Inputs

The program has seven input boxes:



- **Data Notes**
  Use this box for any text notes that you want to record about the data.

- **WQ Objective Conc.**
  The water quality objective or criterion concentration to be assessed. Any real positive number. For CA Ocean discharges use C_o obtained from Table B of the CA Ocean Plan.

- **Background Conc.**
  The ambient pollutant concentration in similar units as the WQ Objective. Any real positive number. For CA Ocean discharges use C_s obtained from Table C of the CA Ocean Plan.

- **Dilution Ratio**
  The ratio of ambient water volume to effluent volume. Any real positive number. For CA Ocean discharges use D_m, the minimum initial dilution as defined in Appendix I of the CA Ocean Plan. No dilution credit is indicated by a zero in this box.

- **RP Percentile**
  The population percentile or quantile to be estimated and compared to the WQ Objective. Any number greater than or equal to 50 or less than 100. The default Ocean Plan value is 95 in order to estimate the 95^{th} percentile of the population distribution.

- **RP Confidence Level**
  The desired confidence level for the one-sided UCB. Any number greater than or equal to 50 or less than 100. The default Ocean Plan value is 95 in order to estimate the upper bound of the selected RP percentile with 95% confidence.

- **Data Set**
  Enter the effluent concentration data, one measurement per line, in the same units as the WQ Objective. Censored data or non-detects are preceded by the "<" symbol. All data values must be real numbers greater than zero.

Other Ways to Enter Input Data

--*Loading Data Using the Windows Clipboard*. Using the menu, click on "File, Load data from Clipboard" to automatically load data currently in the Windows clipboard into the input boxes. You can use any Windows program to load the Windows clipboard. For example, in Excel you simply copy a column of numbers corresponding to the seven input boxes. The first six rows in the column will go into the first six input boxes. The Data Set box will contain the rest of the selected column.

--*Loading Data from a Text File.* Using the menu, click on "File, Load data from File." The program looks for text file names with the *.rpf extension. However, you can import a text file of any extension by selecting "All files (*.*)" from the select file dialog menu. A valid text file contains a column of numbers corresponding to the seven input boxes.

Saving Input Data

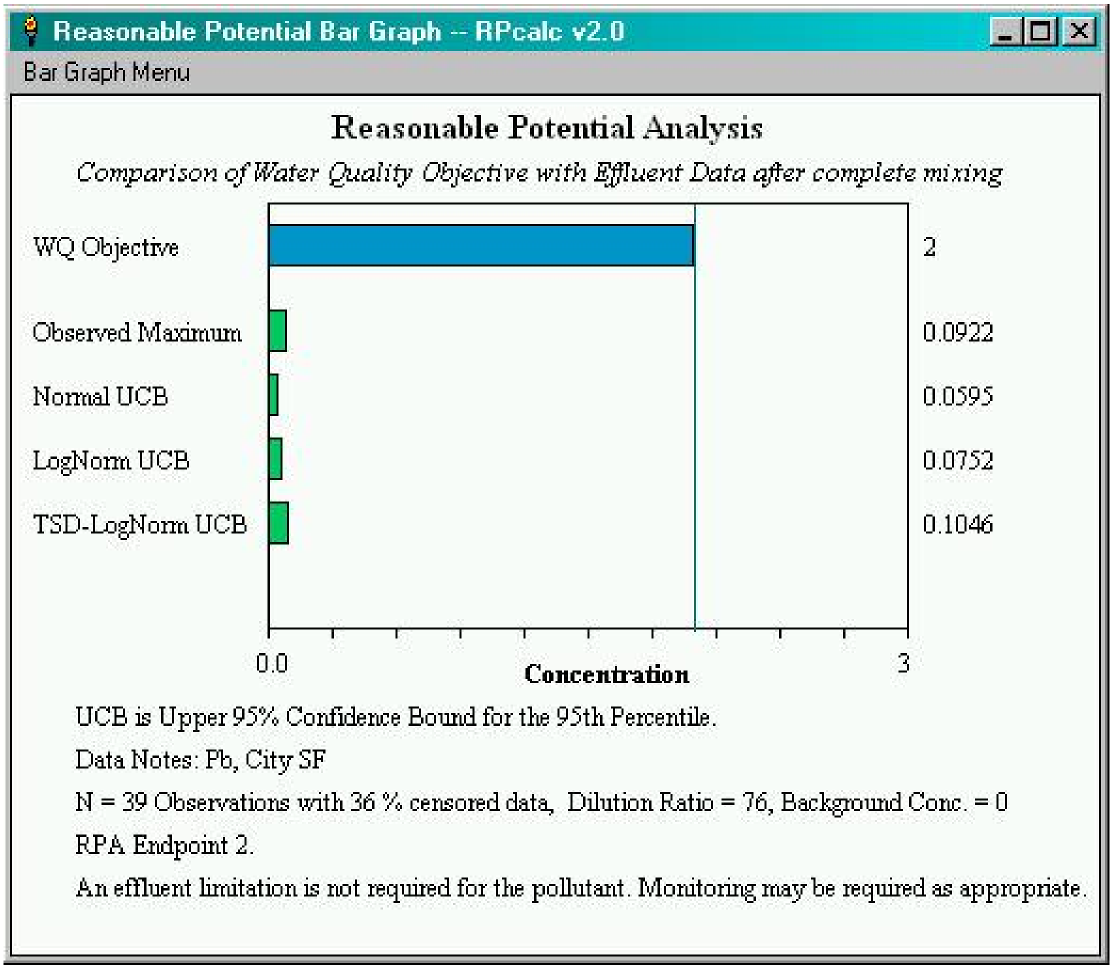
Once data is entered into the seven input boxes, you may save your data in an RPcalc data file by using the menu and clicking on "File, Save current data as…" The data will be saved in a text file with an *.rpd extension.

Performing the Reasonable Potential Analysis

Once you have filled in the input boxes, you may click the "Calculate RP" command button in the lower right hand corner of the main screen to perform the reasonable potential analysis. Alternatively, you may use the program menu by selecting "Data, Calculate RP." The program performs the mathematical calculations according to the flow chart and procedure in Appendix VI of the Ocean Plan (SWRCB 2005). You may use the menu to see the step-by-step

Reasonable Potential analysis procedure and the procedural flow chart by clicking on "Tools, View RPA flowchart."

Once the reasonable potential analysis is performed, the program asks if you want to see the RP calculations outputs. Here is an example of the Bar Graph output:



A more detailed analysis will pop up in a Notepad text file called "RPcalout.txt." Rename this file if you want to save the results. Here is an example of RPcalc's detailed output:

```
*** Reasonable Potential Calculator Output, RPcalc v2.0 ***
  Apr 26, 2005  14:39:53

Inputs:
  1. Data Notes               : Pb, City SF
  2. WQ Objective Conc. (Co)  : 2
  3. Background Conc. (Cs)    : 0
  4. Dilution Ratio (Dm)      : 76
  5. Percentile               : 95
  6. Confidence Level         : 95

After Mixing Data Summary:
                    __N__       __%__       __Min__       __Max__
  Censored Data       14,     35.8974,       0.0130,       0.0649
```

```
  Detected Data       25,     64.1026,      0.0156,      0.0922
  Total               39

Reasonable Potential Analysis Summary:
  RPA Endpoint 2.
    An effluent limitation is not required for the pollutant. Monitoring may be required as
appropriate.
  Rationale:
    Parametric RPA found the LogNormal UCB(.95,.95,39) of 0.0752 does not exceed the Co of 2.

Parametric RPA Details:

Summary Statistics for X-new:
  __N__        __Mean__     __SDev__     __Min__      __Max__      __CV__
     39,      0.0237,      0.0167,      0.0076,      0.0922,      0.7053

Sample Percentiles for X-new:
     __P10__      __P25__      __Median__   __P75__      __P90__      __P95__
      0.0094,      0.0156,      0.0187,      0.0286,      0.0364,      0.0727

Summary Statistics for Ln(X-new):
  __N__        __Mean__     __SDev__     __Min__      __Max__      __CV__
     39,     -3.9424,      0.6354,     -4.8846,     -2.3837,     -0.1612

Upper One-sided Confidence Bounds (UCB) after complete mixing:
  Upper 95% confidence bound for the 95th population percentile with N = 39
  Normal Tolerance Factor, g' = 2.133 (Hahn & Meeker 1991, Table A12)

  __Distribution__    __UCB___________________        __Confidence Coeff.__
  Normal            0.0595 = Mean + SDev * g'              0.9500
  LogNormal         0.0752 = EXP(LnMean + LnSDev * g')      0.9500
  TSD-LogNorm       0.1046 = X(39) * 1.134                 0.9500
  D'n-Free          0.0922 = X(39)                         0.8647
    Note: D'n-Free method requires >=59 samples for the desired confidence.
    Note: TSD-LogNorm based on CV = 0.7053.

Detection or Quantitation Limit Thresholds, 4 present:
  Thresholds in data set are 0.0130, 0.0143, 0.0273, 0.0649

Censored Data Anaysis using Regression on Normal Order Statistics (Helsel & Cohn 1988):
       i            X-obs         X-new        Prob       NormZ
  -----------------------------------------------------------------
      1,         <0.0130,      0.0076,      0.0527,     -1.6192
      2,         <0.0130,      0.0094,      0.1054,     -1.2513
      3,         <0.0130,      0.0109,      0.1581,     -1.0022
      4,         <0.0143,      0.0076,      0.0527,     -1.6192
      5,         <0.0143,      0.0094,      0.1054,     -1.2513
      6,         <0.0143,      0.0109,      0.1581,     -1.0022
      7,          0.0156,      0.0156,      0.2438,     -0.6942
      8,          0.0156,      0.0156,      0.2767,     -0.5926
      9,          0.0169,      0.0169,      0.3097,     -0.4968
     10,          0.0169,      0.0169,      0.3426,     -0.4054
     11,          0.0169,      0.0169,      0.3755,     -0.3172
     12,          0.0182,      0.0182,      0.4085,     -0.2315
     13,          0.0182,      0.0182,      0.4414,     -0.1474
     14,          0.0182,      0.0182,      0.4744,     -0.0643
     15,          0.0182,      0.0182,      0.5073,      0.0183
     16,          0.0208,      0.0208,      0.5402,      0.1010
     17,          0.0208,      0.0208,      0.5732,      0.1845
     18,          0.0208,      0.0208,      0.6061,      0.2692
     19,          0.0221,      0.0221,      0.6391,      0.3560
     20,          0.0260,      0.0260,      0.6720,      0.4455
     21,          0.0260,      0.0260,      0.7050,      0.5387
     22,         <0.0273,      0.0090,      0.0922,     -1.3271
     23,         <0.0273,      0.0116,      0.1845,     -0.8984
     24,         <0.0273,      0.0140,      0.2767,     -0.5926
     25,         <0.0273,      0.0163,      0.3689,     -0.3346
     26,         <0.0273,      0.0187,      0.4612,     -0.0975
     27,         <0.0273,      0.0215,      0.5534,      0.1343
     28,         <0.0273,      0.0248,      0.6457,      0.3736
     29,          0.0286,      0.0286,      0.7613,      0.7105
```

```
     30,          0.0299,      0.0299,      0.7847,      0.7883
     31,          0.0299,      0.0299,      0.8082,      0.8712
     32,          0.0338,      0.0338,      0.8316,      0.9605
     33,          0.0351,      0.0351,      0.8550,      1.0582
     34,          0.0364,      0.0364,      0.8784,      1.1672
     35,          0.0364,      0.0364,      0.9019,      1.2923
     36,          0.0494,      0.0494,      0.9253,      1.4416
     37,         <0.0649,      0.0191,      0.4744,     -0.0643
     38,          0.0727,      0.0727,      0.9658,      1.8225
     39,          0.0922,      0.0922,      0.9829,      2.1178
  ------------------------------------------------------------------
  Linear Regression of (NormZ) vs. (Detected Observations, after transformation)
   __Slope__   __Intercept__  __Correl r__    __N__    __Transformation__
      0.5965,     -3.9187,      0.9715,     25.0000,     Nat. Log
      0.0211,      0.0200,      0.8813,     25.0000,     None
      0.0547,      0.1413,      0.9341,     25.0000,     Sqr. Root
    Data transformation used in regression was Nat. Log.

Unsorted Input Data, Ce:
2.7, <2.1, 2.3, <2.1, <2.1, <2.1, <2.1, <2.1, <2.1, 2.8, 2.3, 7.1, 5.6, 2.0, <1.1,
2.2, 1.4, 1.4, 1.3, 1.3, 1.6, <1.1, 1.3, 1.6, 1.2, <1.1, 1.2, 1.7, 3.8, 1.4,
1.6, <1, 2.6, 2.0, <1, <1, 1.4, <5, 2.8

Sorted Input Data, Ce:
<1, <1, <1, <1.1, <1.1, <1.1, 1.2, 1.2, 1.3, 1.3, 1.3, 1.4, 1.4, 1.4, 1.4,
1.6, 1.6, 1.6, 1.7, 2, 2, <2.1, <2.1, <2.1, <2.1, <2.1, <2.1, <2.1, 2.2, 2.3,
2.3, 2.6, 2.7, 2.8, 2.8, 3.8, <5, 5.6, 7.1

Sorted Input Data after complete mixing, X-obs = (Ce + Dm * Cs) / (Dm + 1):
<0.0130, <0.0130, <0.0130, <0.0143, <0.0143, <0.0143,  0.0156,  0.0156,  0.0169,  0.0169,
0.0169,  0.0182,  0.0182,  0.0182,  0.0182,
 0.0208,  0.0208,  0.0208,  0.0221,  0.0260,  0.0260, <0.0273, <0.0273, <0.0273, <0.0273,
<0.0273, <0.0273, <0.0273,  0.0286,  0.0299,
 0.0299,  0.0338,  0.0351,  0.0364,  0.0364,  0.0494, <0.0649,  0.0727,  0.0922

References:
  Hahn, GJ and WQ Meeker. 1991.  Statistical Intervals: A Guide for
       Practitioners.  Wiley & Sons, NY.  (See especially Sec. 4.4
       Normal D'n UCB, Sec. 5.2.3 D'n-Free UCB.)
  Helsel, DR and TA Cohn. 1988.  Estimation of Descriptive Statistics
       for Multiply Censored Water Quality Data.  Water Resources
       Research, Vol.24, No.12, pp. 1977-2004
  USEPA. 1991. Technical Support Document of Water Quality-based Toxics
       Control, TSD. Sec. 3.3.2. EPA 505 2-90-001
  Programmed by Steve Saiz, CalEPA, SWRCB, ssaiz@waterboards.ca.gov, April 26, 2005
```

Using the "Data" Menu

This program also includes some tools that can be used to graphically examine real or simulated data sets. The "Data" menu contains four data analysis tools that may be useful in assessing your data: Sort, Simulate, Censor, and Plot.

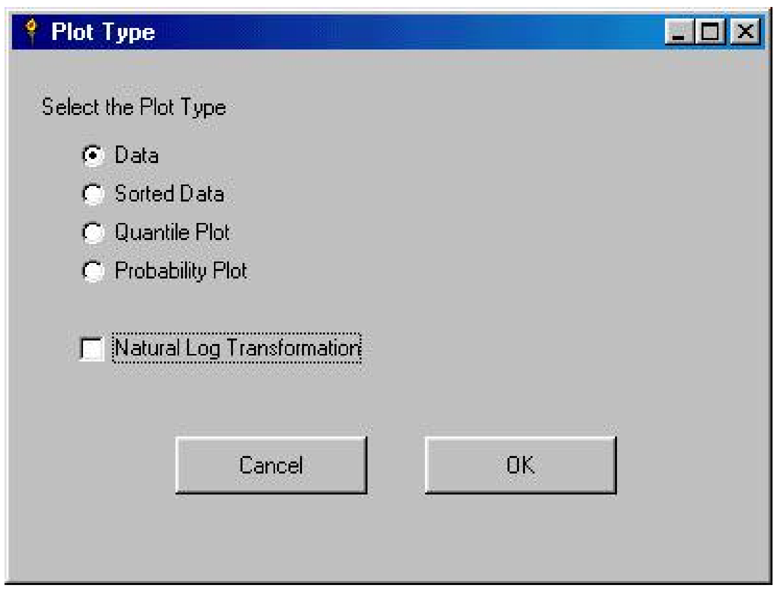
- **Data, Sort**
  This feature will sort the values in the Data Set box from lowest to highest. The "<" symbol is ignored during sorting so that non-detects are sorted according to the detection limit value.

- **Data, Simulate**
  This feature will allow the user to randomly draw samples from a theoretical normal or lognormal distribution and place the random values in the Data Set box. The simulated data set can be analyzed in the usual manner. This feature is useful to perform "what ifs."

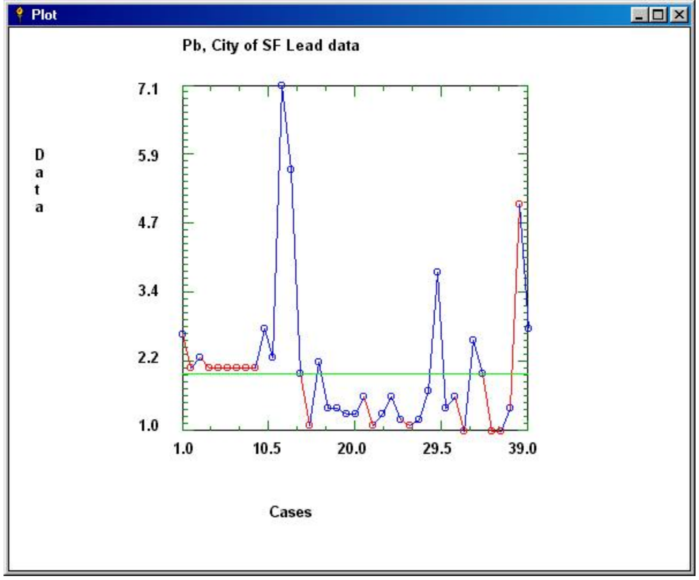
For example, you can see the effect on the UCB of an increased sample size when the mean and standard deviation remain the same.

- **Data, Censor**
  This feature will replace all uncensored data values with a user-supplied limit of detection. For example, if you censor the data set at 5 ug/L then all values less than or equal to 5 will be replaced by "<5" in the Data Set box. Multiple detection limits can be simulated by repeatedly selecting this feature.

- **Data, Plot**
  This feature will allow the user to plot the data currently in the Data Set box. You can perform a logarithmic transformation of the data prior to plotting by checking the "Natural Log Transformation" box. Four types of plots can be made: Data Plot, Sorted Data Plot, Quantile Plot, and Probability Plot.

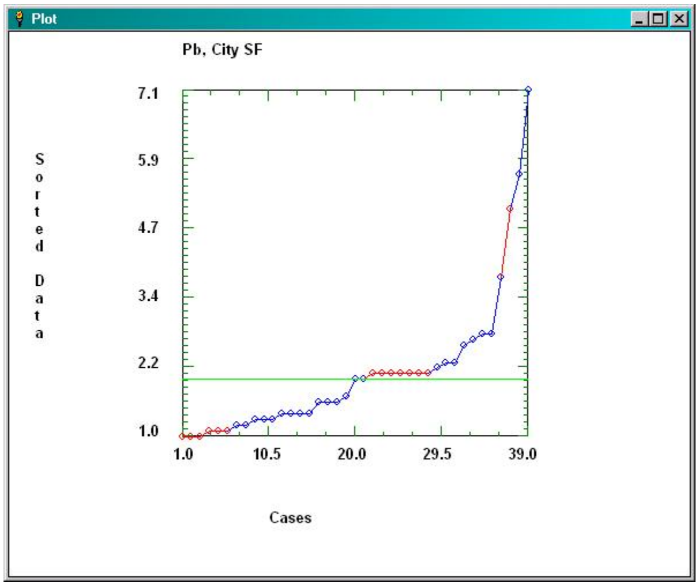


-- ***Data Plot***. This is simply a plot of the data on the Y-axis verses the observed cases on the X-axis. If the data were in chronological order, this would be a time-series plot. Here is an example of a Data Plot:

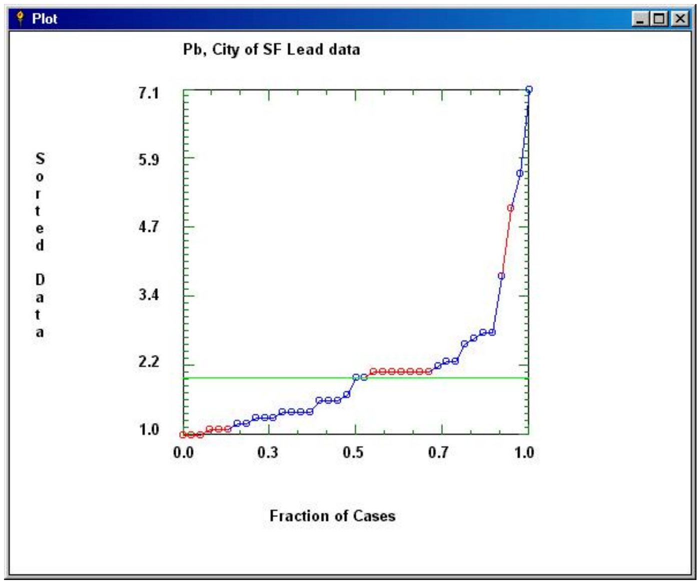


The water quality standard is the green horizontal line. Non-detects are plotted in red and uncensored values are plotted in blue.

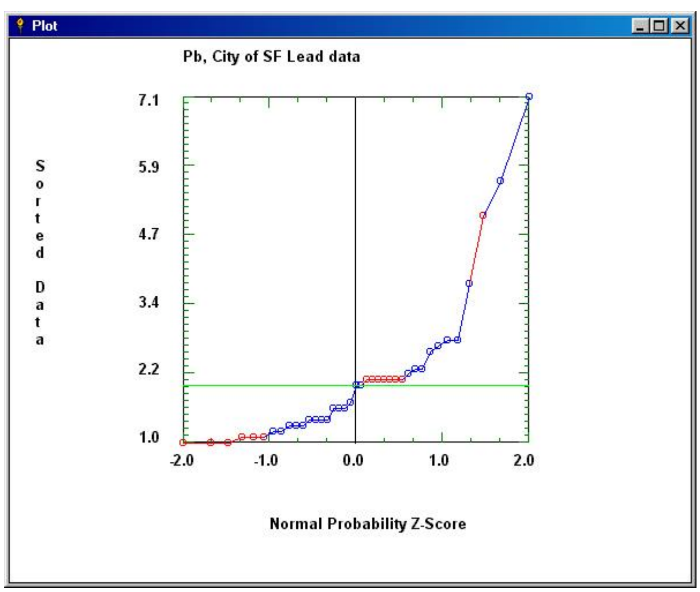
-- *Sorted Data Plot.* This is similar to a Data Plot except the values in the Data Set box are sorted before plotting:



-- *Quantile Plot.* A quantile plot is similar to a Sorted Data plot except the ranked cases are transformed to plotting positions using the Weibull formula, $i/(n+1)$, where i is the ranked case number and n is the sample size. Quantile plots visually portray empirical percentiles of the distribution of the sample data and are also called cumulative frequency plots (Helsel & Hirsch 2001, Sec. 2.1.3):



-- ***Probability Plot*.** A normal probability plot is similar to a quantile plot except the plotting positions are transformed to quantiles of the standard normal distribution, i.e., a normal distribution with a mean of zero and a standard deviation of unity. A normal probability plot is used to visually determine how well your data fit a theoretical normal distribution (Helsel & Hirsch 2001, Sec. 2.1.5). Normally distributed data will appear as a straight line on a normal probability plot. Log-normally distributed data will appear as a straight line on a normal probability plot when the "Natural Log Transformation" box is checked.



Notes on the Upper Confidence Bound Calculations

The output file results will include an estimate of the one-sided, upper $c100$ percent confidence bound for the $p100$th percentile (UCB) after complete mixing under various distributional assumptions: a normal distribution, a lognormal distribution, and any distribution (a distribution free assumption). These UCB estimates are calculated according to the methods in "Statistical Intervals" by Hahn & Meeker (1991) and are equivalent to one-sided, upper tolerance interval bounds. Normal and lognormal tolerance bounds were developed for quality control situations during the 1950's and are based on the statistical non-central *t*-distribution. Distribution free tolerance bounds use the ranked data values.

In addition, an UCB is calculated using the USEPA (1991) Technical Support Document for Water Quality-based Toxics Control method (TSD-LogNorm), which multiplies the sample maximum by a reasonable potential multiplying factor under a lognormal distributional assumption. The TSD-LogNorm UCB can be characterized as a semi-parametric procedure. USEPA guidance on effluent characterization and reasonable potential analyses can be found in Chapter 3 of the Technical Support Document.

A fully parametric analogue of the TSD-LogNorm procedure is possible. The Upper Confidence Bound for a population percentile when data are Lognormally distributed (UCBL) is calculated using the following equation:

$$UCBL_{(c,\ p)} = \exp[\ M_L + S_L\ g'_{(c,\ p,\ n)}\ ],$$

where M_L and S_L are the sample mean and standard deviation of the natural log transformed data and $g'_{(c,\ p,\ n)}$ is a normal distribution tolerance factor for the $p100$th percentile calculated with $c100\%$ confidence for a sample of size n. Hahn & Meeker (1991) provide extensive tables of normal tolerance factors. RPcalc will generate, on the fly, the appropriate $g'_{(c,\ p,\ n)}$ value based on the user inputs. A table of normal tolerance factors for estimating the one-sided, upper 95% confidence bound for the 95^{th} percentile is in the flow chart and procedure in Appendix VI of the Ocean Plan (SWRCB 2005).

Notes on the Analysis of Censored Data

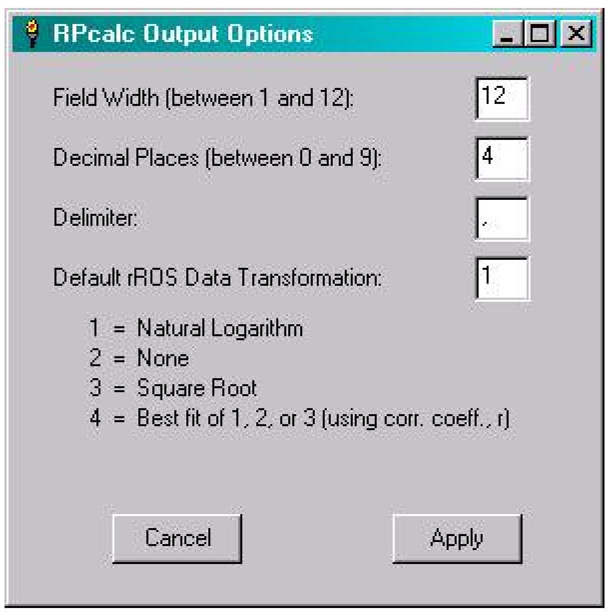
A reasonable potential analysis will be complicated by the presence of monitoring data below the analytical detection limit. Such data are *censored* by a limit of detection or by a limit of quantification, or both, usually on the left tail of the population distribution. Sample results below the limit of detection (i.e., the USEPA Method Detection Limit) are *non-detects*, ND. Monitoring samples at or above the limit of detection but below the limit of quantification (i.e., the Ocean Plan Minimum Level) are *detected but not quantified*, DNQ. Various combinations of data types (NDs, DNQs, or quantified) are theoretically possible depending on the effluent distribution, the limit of detection, and the limit of quantification.

Many statistical techniques exist to account for censored data. In the past, censored data has simply been substituted with zero or one-half the detection limit. These substitution schemes often introduce a bias in the resulting statistics because the detection limit does not accurately represent the effluent concentration in the sample. Because of this, the use of simple substitution methods should be discouraged in favor of better methods.

RPcalc uses the robust *Regression on Order Statistics* technique (rROS) for analyzing censored data because it is robust, unbiased, and has a smaller variance than most other statistical techniques under the lognormal distribution. In addition, this technique will accept multiple detection limits or censoring levels (Helsel & Cohn 1988). The technique uses the uncensored fraction of the data in a probability plot to statistically reconstruct the censored values. A new data set (labeled "X-new" in the output file) is then created containing a combination of uncensored values and reconstructed or "fill in" values derived from the probability associated with the censored data values. Summary statistics (sample mean and sample standard deviation) are then estimated using the new data set. The recently published book, "Non-detects and Data Analysis" by D. Helsel (2005) is also a good reference for the rROS technique and other ways to analyze censored data.

Output Options

Using the menu, click on "Tools, RPcalc Output Options>" to change the current output options.



Field width is the total number of digits, including decimals, used when printing numeric output. *Decimal places* is the number of decimals used in numeric output. *Delimite*r is a character used to separate numbers in the output. Using a comma as a delimiter will allow exporting RPcalc output into Excel or other programs that import comma separated value (CSV) files.

Default rROS Data Transformation defines the default data transformation to be used with the Helsel & Cohn (1988) censored data analysis technique. For most water quality data sets, the *Natural Logarithm* transformation (option 1) is recommended. Using the *Best fit* feature (option 4) will automatically select the data transformation having the highest correlation coefficient (r) in a censored data probability plot amongst the three candidate transformations (natural logarithm, none, or square root). This is an adaptation of the suggestion by Shumway et al. (2002) for improving the rROS technique. Standard RPcalc output shows the regression equations and correlation coefficients for all three candidate transformations.

References

Hahn, G.J. and W.Q. Meeker. 1991. *Statistical Intervals: A Guide for Practitioners.* Wiley & Sons, NY.

Helsel, D. R. 2005. Non-detects and data analysis. Statistics for censored environmental data. Wiley & Sons, NY. http://www.amazon.com/exec/obidos/tg/detail/-/0471671738/ref=cm_rv_thx_view/104-0309763-4903921?%5Fencoding=UTF8&v=glance

Helsel, D. R. and T.A. Cohn. 1988. *Estimation of Descriptive Statistics for Multiply Censored Water Quality Data.* Water Resources Research, Vol.24, No.12, pp. 1977-2004.

Helsel, D.R. and R.M. Hirsch. 2001. *Statistical Methods in Water Resources. Techniques of Water-Resources Investigations of the United States Geological Survey*, Book 4, Hydrologic Analysis and Interpretation, Chapter A3. http://water.usgs.gov/pubs/twri/twri4a3/

Shumway, H. R., R. S. Azari, and M. Kayhanian. 2002. *Statistical approaches to estimating mean water quality concentrations with detection limits*. Environmental Science and Technology 36: 3345-3353.

SWRCB. 2001. *California Ocean Plan. Water Quality Control Plan for Ocean Waters of California*. California Environmental Protection Agency, State Water Resources Control Board. http://www.swrcb.ca.gov/plnspols/oplans/docs/cop2001.pdf

SWRCB 2005. Final Functional Equivalent Document. Amendment of the Water Quality Control Plan for Ocean Waters of California. Issue 1: Reasonable Potential: Determining when California Ocean Plan Water Quality-based Effluent Limitations are Required. April 2005. http://www.swrcb.ca.gov/plnspols/oplans/docs/draft_ffed.pdf

USEPA. 1991. *Technical Support Document for Water Quality-based Toxics Control. Office of Water*. EPA/505/2-90-001

RPhelp20.doc